DOCUMENT RESUME

ED 191 909 TM 800 533

AUTHOR Smith, Philip L.
TITLE Some Approaches to Determining the Stability of Estimated Variance Components.
INSTITUTION Wisconsin Univ., Madison. Dept. of Educational Psychology.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Apr 80
GRANT NIE-G-79-0129
NOTE 25p.; Paper presented at the Annual Meeting of the American Educational Research Association (64th, Boston, MA, April 7-11, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Analysis of Variance; Error of Measurement; High Schools; *Measurement Techniques; Predictive Measurement; Predictive Validity; Reading Tests; *Research Design; *Sampling; Scores
IDENTIFIERS Confidence Intervals (Statistics); *G Analysis; *Generalizability Theory; Jackknifing Technique; Monte Carlo Methods; Variance (Statistical)

ABSTRACT

Accurate estimation of variance components used in generalizability theory is essential for the theory to be viewed as an efficacious mechanism for studying the reliability and validity of a measurement procedure. This paper explores two alternatives for dealing with the apparent instability of small sample size used in determining the accuracy of the variance component estimates resulting from a G study design. First, the accuracy of a confidence interval approach as an indicator of the magnitude of error in variance component estimates was explored. Monte Carlo methods were used to test the adequacy of this approach. Generally, the specific confidence interval procedure used resulted in moderately accurate confidence intervals. Second, the use of sequential designs to provide variance component estimates for multifaceted G studies was explored. The use of sequential designs requires that the investigator set up a series of small G studies designed to produce required variance component estimates and then combine the results of these studies in setting up a more complex D study. This approach was shown to be particularly useful in reducing dispersion effects which can significantly affect least squares estimates of variance components in designs that contain several facets. (Author/RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED191909

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Some Approaches to Determining the Stability of Estimated Variance Components

Philip L. Smith

University of Wisconsin-Milwaukee

"PERMISSION TO REPRODUCE THIS MATERIAL [redacted] HAS BEEN GRANTED BY
P. Smith

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

* Paper presented as a part of a symposium on "Estimating and Interpreting Variance Components in Generalizability Theory" at the meeting of the American Educational Research association, Boston, Massachusetts, April 9, 1980. The preparation of this paper was supported in part by the National Institute of Education (NIE-G-79-0129). However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement of the National Institute of Education should be inferred.

TM 800533

Abstract

Whether or not generalizability theory provides an efficacious mechanism for studying the reliability or validity of a measurement procedure depends upon the accuracy of the variance component estimates that result from a G study design. Previous studies have shown that the accuracy of these component estimates is a function of the sample size used and the complexity of the G study design. Specifically, when small samples are used, the accuracy of variance component estimates computed using least squares procedures is open to considerable question. This is particularly true for complex designs such as multifaceted G studies. The present study explores two alternatives for dealing with this apparent small sample instability.

First, the accuracy of a confidence interval approach as an indicator of the magnitude of error in variance component estimates was explored. Monte Carlo methods were used to test the adequacy of this approach. In general, the specific confidence interval procedure used resulted in moderately accurate confidence intervals.

Second, the use of sequential designs to provide variance component estimates for multifaceted G studies was explored. The use of sequential designs requires that the investigator set up a series of small G studies designed to produce required variance component estimates and then combine the results of these studies in setting up a more complex D study. This approach was shown to be particularly useful in reducing dispersion effects which can significantly affect least squares estimates of variance components in designs that contain several facets.

Introduction

While the number of applications of generalizability theory in the literature has expanded over the past several years, our knowledge of the behavior of generalizability theory and its resulting variance component estimates and coefficient estimates has not expanded greatly over this same period. Recent applications and their interpretations have led to a greater understanding of the usefulness and flexibility of the theory in dealing with measurement problems in general. Unfortunately, many of the problems and limitations associated with the theory that were cited by Cronbach and his associates a number of years ago (Cronbach, Gleser, Nanda & Rajaratnam, 1972) have remained unsolved and have led to a practice where unverified procedures remain the status quo for application. Most notable among these practices ar the artificial balancing of designs that are inherently unbalanced, the casual regard for small sample sizes and the sampling errors that invariably accompany them, and the treatment of negative component estimates as zero in the population. The result is that a useful and flexible procedure and perspective on measurement is being used in a fashion that may be producing unpredictable results and therefore erroneous interpretations. This paper focuses on one area that is particularly important to the conclusions drawn from an application of generalizability theory; some suggested procedures which may allow greater confidence to be placed in the resulting variance component estimates. Because verified and well known procedures

for deriving these estimates from unbalanced designs are unknown, the present paper focuses on procedures that are applicable to balanced designs.

The accurate estimation of variance components for use in generalizability theory is not a trivial matter, but is essential if the theory is to be viewed as an efficacious mechanism for studying the reliability and validity of a measurement procedure. Given G study constraints that provide no hint as to the magnitude of errors that may be present in the resulting variance component estimates, there are at least two intriguing approachs to the accuracy problem. One approach would be to use confidence intervals for variance component estimates which, while not directly helping to provide more accurate estimates, would at least provide some notion as to the stability of the estimates and their resulting interpretations. Such an approach might also be of value if some rough knowledge of the relative magnitudes of variance components and available resources were present. Another approach, one that can be expected to produce more stable estimates while maintaining more reasonable design parameters is the use of sequential designs in G studies (Smith, 1978) where several small G study designs are used as a substitute for one large G study design to approximate variance component estimates for use in D studies.

Previous study (Smith, 1978; Woodward & Joe, 1973) has shown that sampling errors of variance components are dependent, for the most part, on four specific design characteristics: the

sample size used; the complexity of the expressions for the expected mean squares and accompanying computation of variance component estimates; the design configuration (e.g., nesting versus crossing); and the relative magnitudes of the population values of variance components. The relationships between these factors is by no means simple; but in general the larger the sample size and less complex (i.e., fewer facets and less nesting) the design the more stable the estimates are likely to be. Of course, the most desirable solution to the sampling error problem would be to increase the sample size, but practical restrictions and cost consideration often render such a solution impossible. Given such restrictions, the only option available to the investigator is to reduce design complexity. Although some a priori guidelines for this option are desirable and have been suggested (see Smith, 1978), clearly it is important to get some feeling for the amount of error that is present after the data have been collected. Regardless of prior decisions, this step should be considered essential to the conclusions formed about the measurement procedure under study. An obvious approach to this problem would be the use of confidence intervals for variance component estimates.

Confidence Intervals for Variance Component Estimates

The sampling variance of variance component estimates for given sample sizes in balanced ANOVA designs is predictable given the population values of those components (Scheffe, 1959; Graybill, 1961; Smith, 1978). While several methods of

establishing confidence intervals around variance component estimates have been proposed in the literature (Thompson, 1962; Herbach, 1959; Hartley & Rao, 1967; Wang, 1967, Klotz, Milton, & Zachs, 1969; Henderson, 1953; Searle & Henderson, 1961; Yates, 1934 among others) many of these apply to only a few specific design configurations and are therefore not useful in a general sense. An adequate review of these techniques is beyond the scope of this paper. However, there are a few techniques which have been proposed which appear to be generalizable to any anova design configuration that involves the assumptions of generalizability theory.

One technique that appears to have considerable promise for use in generalizability theory is the jackknife procedure proposed by Mosteller and Tukey (1968). Miller (1968) and Collins (1970) have shown this procedure to be reasonably robust an useful for determining rough confidence intervals around variance component estimates. One problem, and one that should not be used to exclude its use in generalizability theory, is the computational labor involved in arriving at jackknifed confidence intervals. Miller (1968), Collins (1970), and Cronbach et al. (1972) all speak to the cumbersone nature of the use of the jackknife procedure. Clearly, a more desirable solution would be one that provides adequate intervals and involves a less cumbersome computing algorithm.

One such technique, apparently effective only for components computed in the form

$$\hat{\sigma}^2_i = [MS_1 - MS_2]/n \qquad (1)$$

has been suggested by Scheffé (1959). Although the procedure has been shown to be useful and fairly accurate for component estimates computed in this form, the procedure is of limited usefulness for most generalizability designs that include more than a single facet.

Satterthwaite (1946) and later Welch (1956) and Graybill (1961) provide a possible solution to the confidence interval problem for any component estimate computed as a weighted combination of mean squares as is the case when least squares procedures are used in generalizability studies. These authors conclude that the actual distribution of variance component estimates computed through a weighted linear combination of mean squares is too situation specific and complicated for practical use. Instead, Satterthwaite suggests that a suitable approximation can be based upon the Chi-square distribution. Welch (1956) has suggested that when estimates take on only positive values with a high probability that a Pearson Type III distribution might be a more suitable approximation and provides the appropriate transformation of the Chi-square to achieve confidence intervals based upon this distribution. Although the Welch solution still contains the restriction that the designs be balanced and random errors normally distributed, it is suggested that the

Welch solution provides better estimates for small samples than does the Chi-square approximation. It should be noted at the outset that any distribution that does not accept non-positive values is not likely to produce adequate confidence intervals where a high probability of negative estimates exists. Previous study has shown this probability to be low only when the number of levels of the facet of interest is adequately represented (say $n \geq 5$) and when the component of interest is relatively large in the population compared to the other components contained in the EMS for that effect (Smith, 1978; Leone & Nelson, 1966). Although somewhat restrictive, investigation of the technique seems warranted.

Given a parameter X that may be expressed as a linear combination of variances (as a variance component estimate from a combination of mean squares),

$$X = h_1\sigma_1^2 + h_2\sigma_2^2 + \ldots + h_i\sigma_i^2 \qquad (2)$$

where X is greater than zero and the h's are known constants (in our case ± 1), we can define

$$n = \frac{(\Sigma\, h_i\sigma_i^2)^2}{\Sigma\, (h_i^2\sigma_i^4/n_i)} \qquad (3)$$

where the n_i are the degrees of freedom associated with the variance (σ^2) and n is the degrees of freedom for a Chi-square distribution which may be used to approximate the distribution of X. Generally the σ_i^2's are unknown and therefore must be replaced by their estimates, in our case the observed mean squares. Thus,

using a χ^2 approximation, a confidence interval for X would take the form

$$P[n\hat{X}/{}_n\chi^2_{\alpha/2} < X < n\hat{X}/{}_n\chi^2_{1-\alpha/2}] = 1 - \alpha \ . \qquad (4)$$

An interval based upon the Pearson type III distribution may be formed by using the following adjustments to the above interval:

$$P[n\hat{X}/A_{\alpha/2} < X < n\hat{X}/A_{1-\alpha/2}] = 1 - \alpha$$

where

$$A_\beta = {}_n\chi^2_\beta - C$$

$$C_\beta = 2/3\ (2Z^2_\beta + 1)(Xs^{-2}t - 1)$$

$$t = \Sigma h_i \sigma_i^6 / n_i^2$$

$$s = \Sigma h_i^2 \sigma_i^4 / n_i$$

where Z_β is the normal deviate corresponding to the β percentage point.

To evaluate this technique, five two-facet design configuration (I and J where P is the face of differentiation) were simulated using monte carlo techniques. In each design, the components corresponding to the P, I, J, IJ, PI, PJ, and Residual effects were represented by the values 20, 12, 47, 23, 35, 64, and 76 respectively. These values were chosen so that a) sufficient variance would exist between these population values and b) so that the total variance was proportioned in a way that resembled estimates obtained in certain past applications of the two facet design. The five design configurations corresponded to one completely crossed design, one in which I was nested within

J, one in which J was nested within I, one in which I was nested within P, and one in which J was nested within P. Variations in nesting were used because of the anticipated dispersion effects created by this nesting (see Leon & Nelson, 1966; Smith, 1978). The number of levels of each facet in each configuration was varied at 2, 4, 8, and the number of levels of the face of differentiation (P) was varied at 25, 50, 100. The resulting 27 variations of each design type were simulated with between 300 and 500 replications each. Because of the high incidence of negative component estimates, only the largest 12 variations of each design could be retained for use with the confidence interval approach. It should be noted at this point that the smallest of the designs retained contained 400 observations, which may be thought of as a minimum. Even at that, the intervals for the I and J component could not be reliably estimated because of the small estimated degrees of freedom (see equation 3) associated with these estimates. This suggests at the outset that if the technique is to be used reliably, the number of levels of each of the facets of interest needs to be adequately represented, with at least 10 levels of any facet for which the interval is desired represented in the study. This is not encouraging given the magnitude of most designs seen in the literature.

Table 1 shows the nominal and observed proportion of intervals that did not include the population value (α = .05 in this case) collapsed acress design variation within design type. The collapsing was done because the degrees of freedom computed for

this technique is variable dependent and therefore any comparison across degree of freedom categories is likely to be misleading. In general, as the mean square estimates diverge from their expected values, so will the computed degrees of freedom for the test. Because of this, it can be expected that for a given component and design, the confidence interval will have a decreasing probability of including the population value as the degrees of freedom diverge from its expected value. The degrees of freedom are dependent on the number of levels of the other facets included in the design as well. In complex expressions for computing component estimates this effect is also complex and therefore may be misleading.

Insert Table 1 here

The data in Table 1 suggest that for components computed as the difference between two mean squares and for $\sigma^2(\delta)$ which in this case contains components computed in this form, the technique produces reasonable estimates. For terms computed in a more complex fashion, the interval is somewhat liberal, particularly in the upper limit. Two tentative conclusions may be drawn from these data. First, in order to use the techniques with any accuracy, the probability of obtaining negative estimates should be small. This condition is usually met when design dimensions are large (at least 10 levels of each facet) and when the component of interest is reasonably large with respect to the other

components involved in the estimation. Second, the degrees of freedom computed for the interval should be reasonably large (at least 10). This condition is met again only when design dimensions are reasonably large. The technique appears to be most suitable for components that are computed as the weighted difference between two mean squares. However, a simpler, and perhaps more accurate, solution exists for this circumstance (see Scheffe, 1959; and pg. 5).

The use of sequential designs

The above discussion is discouraging since it appears that the confidence interval approach is most useful when circumstances indicate that they are least needed. One rule of thumb in the use of variance component estimates would be that if the interval can't be formed due to the restrictions cited above, one can expect estimates to be misleading and of little value anyway. So, what are the alternatives? The most obvious is the use of large enough samples to produce stable estimates, but this is often impractical or in some cases impossible. Another viable alternative may be the use of sequential designs for the estimation problem.

The emphasis of a G study is on accurate estimates of variance components. Subsequent use of these estimates (D studies) bear no particular dependence upon how data are collected in G studies as evidenced by the flexibility in D study designs afforded by the G study estimates. As a matter of fact, one extreme solution to the instability of G study estimates would be

not to carry out a G study at all, but rather use some intuitive notion regarding the relative magnitudes of variance components in setting up the design of D studies. This procedure is implicit in the way that measures are used under the classical theory (e.g., situation variance is assumed to be so small as to not influence the measurement results). In some cases, this intuition may be of more value than actual G study estimates. However, empirical verification is still desirable, thus G study estimates are still the goal for D study application.

Since sampling errors of variance component estimates are a function of the design configuration and the expressions for expected mean squares, thus the procedures used to compute the estimates, one effective solution to the sampling error problem might be to alter these factors in a G study design to obtain more accurate component estimates. The sampling theory of variance component estimation indicates that greater stability can be obtained if fewer variance terms enter the expected mean square for the source of interest and/or if the number of mean square estimates used to compute the component estimate is reduced. In general, the latter strategy will lead to the greater pay off in terms of accuracy. The goal of a procedure that takes advantage of these factors would be to simplify expressions for expected mean squares and/or simplify the computing algorithms for variance component estimation by altering the complexity and configuration of the G study design. Some loss of information regarding specific components can be expected from

such a procedure: however, a sequence of such designs could serve the same purpose while minimizing this loss of information. In addition, such a procedure could lead to a savings of the total number of observations required to obtain estimates within a specified range of accuracy. For example, in a two facet fully crossed design, adequate variance component estimates probably could be obtained if 10 levels of each face were sampled, a total of 1000 observations. Theoretically, the only reason to go to a design of such magnitude would be to provide adequate estimates of the components of each face, since because of their more simple nature, stable interaction components could be obtained using a much smaller sample size. As an alternative to the larger design, one could set up three smaller designs: one which has P crossed with I, another with P crossed with J, and the third a mini-replication of the desired full design except where the number of levels of each facet were say, 6. Using this procedure, all of the desired components could be extracted for use in D studies with a savings of over half the observations required of the larger designs. Since the P, I, and J components are estimated in less complex fashion in the first two designs, one could expect these estimates to be equally or more stable than those based upon the larger design.

As an example, consider the analyses reported in Table 2. The data used for these analyses were standardized test score results for high school sophomores using the reading subtest of the Metropolitan Achievement Test. For illustrative purposes,

students were nested within schools and items were grouped according to type within test strata. The design is represented by four facets with students nested within schools which are crossed with items nested within strata. Although the nature of these particular data are such that large samples might be reasonable in an applied setting, the principle to be extracted is for those types of data where such large samples would not be readily available.

Insert Table 2 about here

The first column in Table 2 indicates the sources of variation derived when the full design was analyzed. The second column shows the variance component estimates for these sources in a design application where the number of schools (n_s) equals 14, the number of pupils (n_p) equals 100, the number of strata (n_f) equals 10, and the number of items (n_i) equals 6 (a total of 84,000 observations). The values contained in this column can probably be considered reasonably accurate estimates of the corresponding population values. However, a design of this magnitude is usually impractical for many applications of generalizability theory. A more reasonable design of the same configuration would be one where 50 students each from seven schools are observed on six items in each of five test strata (a total of 7,350 observations). The third column of Table 2 shows

the largest and smallest estimated values of each of the components generated from ten randomly sampled replications of a design of this size. These values indicate considerable sampling fluctuation, particularly for the S, F, and SF component estimates which are computed using four mean squares. The errors that could be made in interpretation are obvious.

There are a number of sequential design strategies that could be used to reduce the number of total observations required to a level comparable to that used in the "small" design while simultaneously providing what can be expected to be more stable component estimates. One such plan would require two sequential designs: one in which a single school would be given a large proportion of the total exam (say 5 strata or 30 items) and the other in which several schools (say seven for comparability) are given a single strata. With 100 students sampled in each school, the total number of observations required in such a strategy would be 7,200. Although these designs require variance component estimation algorithms similar to the full design, fewer variance terms enter the expressions for the expected mean squares in these designs, thus some gain in accuracy can be expected. Columns 4 and 5 of Table 2 show the highest and lowest value of each of the variance component estimates obtained from ten randomly sampled replications of the sequential design strategy outlined above. Notice that the high to low intervals produced by this strategy are in general shorter than those obtained when the full design was used even though the number of

levels of the facets of interest are, in general, comparable. Note also that this procedures is not without its price; information regarding some specific components is lost, but most of these are components that are confounded in the full design so that there is not complete loss in most situations.

Another sequential strategy, one that would provide estimates of more of the desired components, reduce the number of variance terms entering the expressions for the expected mean squares, and would simplify the computation of the estimation of the two main effects components (that for schools and that for strata; reducing the number of mean squares involved from four to two), is possible. This strategy also requires two designs, one in which subsets of pupils within schools are given different test strata, and the other in which different subsets of items within strata are assigned to different schools. Columns 6 and 7 of Table 2 shows the largest and smallest estimated values for each component from 10 randomly sampled replications of this strategy. In these designs, the number of levels of the school and strata facets were set at values comparable to those used in the previous strategies, as was the total number of observations required (here, 6650). The intervals produced by this strategy are comparable to those produced by the earlier sequential strategy, except the present strategy provides better estimates of the S and F components, particularly when the computing algorithm is simplified (see asterisked items in Table 2). Although again some specific information is lost, the payoff in

accuracy, particularly for this latter strategy, seems to be worth the price.

These results indicate that sequential designs are useful for obtaining more stable estimates of variance components while at the same time resulting in a savings of observations. This can be particularly critical when the procedure under study requires considerable time and money, as with observation schedules and interviews. For example, were "observers" are crossed with "occasions," several observations of each level of the face of differentiation would be required for accurate estimation, which would be both costly and time consuming. It is probably less advantageous for objective tests, as has been used as an example here. In general, the trade-off that is made is a sequence of less complex design with perhaps some loss of specific component estimates for a smaller and thus less expensive and time consuming G study sample.

Conclusion

With respect to accurate variance component estimation, the use of large sample sizes appears the most promising solution. Although some confidence interval techniques, such as the jackknife procedure, produce reasonable intervals, other less cumbersome techniques require the same large sample sizes for practical use. One promising alternative to either of the above solutions is the use of sequential designs for the estimation of variance components. Such a procedure takes advantage of the fact that certain measurement procedures can be construed in a

number of different fashions, some of which may involve less complex expressions for expected mean squares, which can be expected to result in more stable component estimates while at the same time reducing the total number of observations required, perhaps substantially.

References

Collins, J. D. Jackknifing Generalizability. Unpublished doctoral dissertation, University of Colorado, 1970.

Cronbach, L. J., Gleser, G., Nanda, H., & Rajaratnam, N. The dependability of behavioral measurements: Theory of generalizability for scores and profiles. New York: Wiley, 1972.

Graybill, F. A. An introduction to linear statistical models: Volume I. New York: McGraw-Hill, 1961.

Hartley, H. O., & Rao, J. N. K. Maximum likelihood estimation for the mixed analysis of variance model. Biometrika, 1967, 54, 93-108.

Henderson, C. P. Estimation of variance and covariance components. Biometrics, 1953, 9, 226-252.

Herbach, L. H. Properties of model II type analysis of variance tests. Annals of Mathematical Statistics, 1959, 39, 939-959.

Klotz, J. H., Milton, R. C., & Zachs, S. Mean square efficiency of estimates of variance components. Journal of the American Statistical Association, 1969, 64, 1383-1402.

Leone, F. C., & Nelson, L. L. Sampling distributions of variance components I. Empirical studies of balanced nested designs. Technometrics, 1966, 8, 457-468.

Miller, R. J. Jackknifing variances. Annals of Mathematical Statistics, 1968, 39, 567-582.

Mosteller, F., & Tukey, J. W. Data analysis, including statistics. In G. Lindzey & E. Aronson (Eds.), Handbook of Social Psychology: Volume II. Reading, MA: Addison-Wesley, 1968.

Satterthwaite, F. E. An approximate distribution of estimates of variance components. Biometrics, 1946, 2, 110-114.

Scheffé, H. The analysis of variance. New York: Wiley, 1959.

Searle, S. R., & Henderson, C. R. Computing procedures for estimating components of variance in the two way classification mixed model. Biometrics, 1961, 17, 607-616.

Smith, P. L. Sampling errors of variance component estimates in small sample multifacet generalizability studies. Journal of Educational Statistics, 1978, 3, 319-346.

Thompson, W. A. The problem of negative estimates of variance components. Annals of Mathematical Statistics, 1962, 33, 273-289.

Wang, Y. Y. A comparison of several variance component estimators. Biometrika, 1967, 54, 301-305.

Welch, B. L. On linear combinations of several variances. Journal of the American Statistical Association, 1956, 51, 132-148.

Woodward, J. A., & Joc, G. W. Maximizing the coefficient of generalizability in multi-facet decision studies. Psychometrika, 1973, 38, 173-181.

Yates, F. The analysis of multiple classification with unequal numbers in the different classes. Journal of the American Statistical Association, 1934, 29, 51-66.

Table 1

Confidence Intervals for Selected Variance Components

in Five Design Configurations

DESIGN CONFIGURATION	SOURCE	N	Proportion Below .975 Nominal	Proportion Below .025 Nominal
P x I x J	P	2674	.872	.058
	IJ	3355	.971	.030
	PI	3599	.964	.019
	PJ	3600	.973	.030
	error (δ)	3600	.971	.022
	error (Δ)	3600	.780	.022
P x (I:J)	P	4131	.910	.035
	J,IJ	4800	.967	.027
	PI	4763	.955	.025
	error (δ)	4800	.971	.017
	error (Δ)	4800	.922	.105
P x (J:I)	P	5222	.916	.034
	I,IJ	5991	.961	.029
	PJ	5951	.951	.025
	error (δ)	6000	.964	.017
	error (Δ)	5867	.816	.145
(J:P) x I	P*	1553	.852	.037
	J,PJ	3600	.966	.023
	PI	3589	.958	.025
	error (δ)	3600	.971	.018
	error (Δ)	3600	.944	.101
(I:P) x J	p*	1650	.852	.030
	I,PI	3600	.973	.027
	PJ	3599	.960	.028
	error (δ)	3600	.972	.018
	error (Δ)	3552	.802	.165

* based on six largest designs only

TABLE 2

Variance component estimates for some selected estimation strategies

DESIGN	LARGE	SMALL		SEQUENTIAL STRATEGY 1				SEQUENTIAL STRATEGY 2			
CONFIG	P:S x I:F	P:S x I:F		P xI:F		P:S x I		P:S,F) x I		P xI:(S,F)	
$n_p/n_s/n_i/n_f$	100/14/6/10	50/7/6/5		100/1/6/5		100/7/6/1		15/7/6/5		50/7/2/5	
COMPONENT		Low	High	Low	High	Low	High	Low	High	Low	High
S	.0123	.0000[a]	.0352	---	---	.0026	.0293	.0092*	.0177*	.0076	.0178
F	.0130	.0000(5)[a]	.0134	.0029	.0351	---	---	.0086	.0177	.0088*	.0171*
P PS	.0307	.0262	.0515	.0138 ---	.0571 ---	.0256	.0401	.0405+	.0519+	.0267	.0351
I IF	.0193	.0104	.0308	.0112	.0264	.0025 ---	.0473 ---	.0153	.0282	.0183+	.0257+
SF	.0016	.0000	.0045	---	---	---	---	.0018	.0055	.0009	.0049
SI SIF	.0015	.0000(2)[a]	.0086	---	---	.0003 ---	.0033 ---	.0001	.0043	.0183+	.0257+
FP FPS	.0175	.0000[a]	.0160	.0079 ---	.0235 ---	---	---	.0405+	.0519+	.0113	.0245
PI PIS PIF PISF E	.1458	.1272	.1635	.1216	.1569	.1123	.1797	.1389	.1550	.1399	.1525

[a] indicates negative value treated as zero. Number in parenthesis indicates number of negative values treated as such.

* computation simplified from four terms to two terms

+ within a column indicates confounding